DATA & RESEARCH DATA MANAGEMENT

WHAT IS DATA?

Data is anything you perform analysis on. Most researchers work with data but sometimes use different terms to describe their data such as research assets or research output. Data varies widely by academic discipline and research project.

A few types of data include:

- Experimental measurements
- Research notes or lab notebooks
- Physical samples
- Images
- Audio or video
- Survey responses

Data can be both digital and physical. For example, you may have online survey responses (digital data) and paper research notes (physical data).

WHAT IS RESEARCH DATA MANAGEMENT (RDM)?

Research data management (RDM) is the compilation of small practices that make your data easier to find, easier to understand, less likely to be lost, and more likely to be usable during a project or ten years later. Additionally, many funding agencies have research data mandates for funded projects.

RDM STEPS & EXAMPLES OF BEST PRACTICES

While RDM is rapidly developing, the individual steps of RDM are well defined (see below). Each step has many best practices, one of which is listed below.

- **Planning**
  - Write a data management plan (DMP) and share it with collaborators

- **Documenting**
  - Write documentation to describe how the data is organized

- **Storing**
  - Back up your data regularly and test that you can access your backed up data

- **Sharing**
  - Retain data that supports publications or can’t easily be reproduced

- **Preserving**
  - Deposit your data into a data repository for long-term preservation

RDM practices are highly dependent on the research project and discipline. Some disciplines have well defined RDM standards and community practices (such as using disciplinary data repositories) while others are still developing theirs. **OU Libraries are here to help you design, develop and implement RDM practices into your research workflows.**

Source: *Data Management for Researchers* by Kristin Briney
WHY IS RDM IMPORTANT?

YOU DON’T WANT TO LOSE YOUR DATA!

SAVE VALUABLE TIME AND RESOURCES

SUPPORT OPEN ACCESS AND OPEN DATA

COMPLY WITH FUNDING AGENCY GUIDELINES

FUNDING AGENCY MANDATES

Most federal funding agencies mandate RDM for funded projects

• Federal funding agencies with RDM mandates include: NIH, NSF, DOD, and DOE
• Their mandates can vary widely so always check their website for the latest guidelines
• The most common mandate is the inclusion of a data management plan (DMP) with new grant proposals

Other research funders are starting to mandate data management

• Examples include The Alfred P. Sloan Foundation and The Bill and Melinda Gates Foundation
• Some academic journals also have data management mandates (such as Science and Nature)

OU LIBRARIES CAN HELP YOU WITH RDM

We can help you with every stage of RDM, from data management planning to preserving your data. We offer RDM consultations (individual or small group) and workshops of RDM topics.

Contact Joanna Thielen, Research Data Librarian, for more information
jthielen@oakland.edu | 248-370-2477

Source: Data Management for Researchers by Kristin Briney
DATA MANAGEMENT PLANS

WHAT IS DATA MANAGEMENT PLAN?
A data management plan (DMP) is a formal document stating your data management practices both during and after a specific research project.

Many funding agencies require a DMP with new grant proposals. In general, a DMP should include the types of data being collected, data formats (including metadata), and your plans for data storage, sharing, and preservation.

BEST PRACTICES

Check funding agency guidelines before writing
• Funding agency guidelines vary widely so always check their website for the latest guidelines before starting to write a DMP

Write a DMP (even if your funding agency doesn't require one)
• Write a separate DMP for each research project
• Be as comprehensive and specific as possible with your data management practices. For example, where will you store your data and how will you back up your data?
• Write a DMP that's ~2 pages

Share and update the DMP throughout the project lifecycle
• Share the DMP with everyone involved in the project
• Put the data management practices in the DMP into practice
• Update the DMP as the project evolves

Source: Data Management for Researchers by Kristin Briney
REASONS TO CREATE A DMP

- SAVES VALUABLE TIME AND RESOURCES
- AVOIDS DUPLICATION OF RESEARCH (E.G. RE-COLLECTING DATA)
- PROVIDES CONTINUITY IF RESEARCHERS JOIN OR LEAVE THE PROJECT
- MANY FUNDERS REQUIRE A DMP WITH GRANT PROPOSALS

RESOURCES FOR WRITING A DMP

University of Minnesota's Funding Agency Guidelines
- URL: bit.ly/UM_agencyguidelines
- Lists requirements for most federal funding agencies (NIH, NASA, DOE, DOT, etc.)

Columbia University's Data Management Planning Self-Assessment
- URL: bit.ly/CU_selfassess
- Extensive list of questions to consider when writing a DMP

ICPSR's Framework for Creating a Data Management Plan
- URL: bit.ly/ICPSR_DMP
- Outlines many of the recommended DMP sections, including example text for each section

DMPTool
- URL: dmptool.org
- Free, online resource that guides you through writing the necessary sections of a DMP
- Extensive repository of sample DMPs

OU LIBRARIES CAN HELP WITH THE STAGES OF WRITING A DMP, FROM FINDING FUNDING AGENCY GUIDELINES TO PROOFREADING.

Contact Joanna Thielen, Research Data Librarian, for assistance
jthielen@oakland.edu | 248-370-2477

Source: Data Management for Researchers by Kristin Briney
WHAT IS DATA DISCOVERY?

Due to the proliferation of digital data and the increase in data sharing mandates, vast amounts of research data can now be discovered online. In most cases, you can re-use this data for your own research.

COMMON DATA SOURCES BY DISCIPLINE

<table>
<thead>
<tr>
<th>Source</th>
<th>Disciplines</th>
<th>URL</th>
<th>Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataverse</td>
<td>All</td>
<td>dataverse.harvard.edu</td>
<td>Open repository of data from all disciplines</td>
</tr>
<tr>
<td>figshare</td>
<td>All</td>
<td>figshare.com/browse</td>
<td>Open repository of data from all disciplines</td>
</tr>
<tr>
<td>Data.gov</td>
<td>All</td>
<td>data.gov</td>
<td>“Home of the U.S. government’s open data”</td>
</tr>
<tr>
<td>ZENODO</td>
<td>All</td>
<td>zenodo.org</td>
<td>Originally focused on the sciences, now contains data from all disciplines</td>
</tr>
<tr>
<td>ICPSR</td>
<td>Social sciences</td>
<td>icpsr.umich.edu</td>
<td>Political and social research; 1960s to present</td>
</tr>
<tr>
<td>iPoll*</td>
<td>Social sciences</td>
<td>bit.ly/iPolldata</td>
<td>US public opinion poll data</td>
</tr>
<tr>
<td>American FactFinder</td>
<td>Social sciences</td>
<td>factfinder.census.gov</td>
<td>Demographic data about US, Puerto Rico and island territories from censuses and other surveys</td>
</tr>
<tr>
<td>Pew Research Center</td>
<td>Social sciences</td>
<td>pewresearch.org</td>
<td>Nonpartisan fact tank that provides public opinion polls and demographic data</td>
</tr>
<tr>
<td>UK Data Archive</td>
<td>Social sciences &amp; humanities</td>
<td>data-archive.ac.uk/find</td>
<td>“UK’s largest collection of digital research data in the social sciences and humanities”</td>
</tr>
<tr>
<td>Dryad</td>
<td>Sciences</td>
<td>datadryad.org</td>
<td>Mostly biology data</td>
</tr>
<tr>
<td>GitHub</td>
<td>Sciences</td>
<td>github.com</td>
<td>Code</td>
</tr>
</tbody>
</table>

Source: *Data Management for Researchers* by Kristin Briney
Databases provided by OU Libraries. Must be current OU student, faculty or staff to access.

**FINDING DATA USED IN A SPECIFIC ARTICLE**

If you’re looking for data used in a specific article, there are several ways to find it:

- **Look for the data as a supplementary file to the article**
  - Look on the publisher’s website to find any supplementary files associated with the article

- **Contact the corresponding author**
  - In your request, be as specific as possible about what data you are requesting
  - Unfortunately, the author may take a while responding to your request and/or ultimately may not be able to provide the data

- **Contact Joanna Thielen, Research Data Librarian**
  - Librarians are experts at finding things!
  - Contact info: jthielen@oakland.edu, 248-370-2477

**HOW TO BE A GOOD DATA CONSUMER**

As a data consumer, you are re-using a dataset that was collected by another researcher. Here are a few simple ways that you can be a courteous consumer of that data:

1. **Look at the license on the dataset** – Not all datasets will have a license on them. Common licenses are a Creative Commons license (creativecommons.org/licenses/) or an Open Data Commons license (opendatacommons.org/licenses/). If the data does have a license, carefully look at the terms of the license to understand how you can (and can’t) re-use the data.

2. **Give credit where credit is due** – In any publication or presentation, give the complete citation for the dataset (like you would for a journal article). At a minimum, the citation should include the authors, date of publication, title, publisher or distributor and electronic location or identifier (such as a permanent URL or DOI). See this online guide for citing datasets using MLA, APA and Chicago styles: bit.ly/dataset_citation

**OU LIBRARIES CAN HELP WITH DATA DISCOVERY**

Contact Joanna Thielen, Research Data Librarian, for assistance
jthielen@oakland.edu | 248-370-2477

Source: *Data Management for Researchers* by Kristin Briney
DATA STORAGE & SECURITY

WHY IS DATA STORAGE & SECURITY IMPORTANT?

Your research data is incredibly valuable so you don’t want to lose it. Data loss can happen due to theft, hardware/software failure, computer virus, natural disaster or accidental loss. To prevent data loss, your data should be stored safely and securely. When planning your data storage and security procedures, include both physical and digital data. A general rule of thumb for data storage is to follow the LOCKSS principle: lots of copies, keep stuff safe.

BEST PRACTICES: GENERAL DATA STORAGE

Practice the 3-2-1 rule
- You should have 3 copies of your data stored on 2 media types with 1 copy offsite
- You may want additional copies if your data can’t be easily reproduced

Back up your data regularly
- The frequency of backups depends on the research project but, in general, data should be backed up weekly
- Test your data back ups regularly to ensure that you can access the backed up data

Document your storage and security procedures
- Share with these procedures with your collaborators
- Practice the 3-2-1 rule for your documentation as well

BEST PRACTICES: PHYSICAL DATA STORAGE & SECURITY

Think critically about storage location and size
- Ideally, the storage location will help prevent data degradation (such as paper deterioration due to sunlight)
- For large amounts of physical data, create a label and organization system

Create digital copies of any physical data
- For example, you can scan your lab notebooks or research notes in order to create digital back ups

Source: Data Management for Researchers by Kristin Briney
STORAGE HARDWARE

The ideal hardware is dependent on the research project, including the quantity of files, types of data and preferred back up method. However, all hardware will eventually fail or become obsolete. As part of your data storage procedures, plan to migrate data to new storage hardware every few years.

<table>
<thead>
<tr>
<th>Hardware</th>
<th>Recommended?</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal computer</td>
<td>YES</td>
<td>Prone to theft or loss</td>
</tr>
<tr>
<td>External hard drive</td>
<td>YES</td>
<td>Subject to degradation; lifetime is ~5 years</td>
</tr>
<tr>
<td>Local server or drive</td>
<td>YES</td>
<td>Limited storage space</td>
</tr>
<tr>
<td>CD/DVD</td>
<td>YES</td>
<td>Subject to degradation due to mishandling; can be laborious to use</td>
</tr>
<tr>
<td>USB flash drive</td>
<td>NO</td>
<td>Easy to lose; very fallible</td>
</tr>
<tr>
<td>Cloud services*</td>
<td>Maybe</td>
<td>Don’t use for confidential data; by agreeing to their Terms of Service, you may be giving the company a license to use your data</td>
</tr>
</tbody>
</table>

* OakShare is the only OU approved cloud service for confidential data storage.

BEST PRACTICES: CONFIDENTIAL DATA STORAGE

According to OU Policy 860, confidential research data includes (but is not limited to) “information related to a forthcoming or pending patent application, grant applications and proposals, and information related to human subjects.” Confidential data needs to be stored more securely than non-confidential data.

Physical confidential data
- Control access to physical storage locations
- Create a log system to track who accesses the data
- Limit transportation to only essential circumstances

Digital confidential data
- OU-approved storage: OakShare
- Password protect files
- Don’t put confidential data on the internet or in email
- Encrypt all files and devices that access confidential data

OU LIBRARIES CAN HELP WITH DATA STORAGE & SECURITY

Contact Joanna Thielen, Research Data Librarian, for assistance
jthielen@oakland.edu | 248-370-2477

Source: Data Management for Researchers by Kristin Briney
DATA DOCUMENTATION & ORGANIZATION

WHY IS DATA DOCUMENTATION & ORGANIZATION IMPORTANT?

A major challenge with research is keeping track of your research data. Data documentation provides context to understand and use your data while data organization provides the structure for your data. Following consistent data organization and documentation procedures will help you (and others) to be able more easily to find, understand and use your research data, whether you need the data tomorrow or in ten years.

Most researchers already have some sort of data documentation and/or organization method. Some common examples are shown below.

**Examples of data documentation methods**
- Research notes or lab notebooks
- Code books
- Readme.txt files

**Examples of data organization methods**
- File naming conventions (FNC)
- Tracking file versions

BEST PRACTICES: DATA DOCUMENTATION

**Create a procedure for creating documentation about your data**
- The type of documentation needed and how to capture it is dependent on the research project
- Share this procedure with collaborators
- Make sure that your documentation is securely stored and backed up along with your research data

**Documentation should provide as much context as possible**
- In general, record the who, what, when, where, why and how relating to the data
- Don't forget to document abbreviations, important names, locations, etc.
- University of Minnesota Library guide on data documentation: bit.ly/data_doc

**Be consistent with your documentation practices**
- Try to automate these practices as much as possible
- Creating documentation can be time consuming but remember that you will likely benefit from detailed documentation months or years after the research project ends

Source: *Data Management for Researchers* by Kristin Briney
DATA ORGANIZATION: FILE NAMING CONVENTIONS (FNC)

Often, researchers keep their files in a folder hierarchy. But it’s unlikely that a file could be correctly identified if that hierarchy disappears. A file naming convention (FNC) describes the contents of the file and its relation to other files, without depending on a folder hierarchy. It usually consists of a string of 4-7 elements (see below).

Example of a FNC

“ymmdc-PPACEN-FIN-BudgetProposal-v1-00.xls”

- A. Reverse Creation Date
- B. Project Name
- C. Dept Section
- D. Doc Name
- E. Version Number
- F. Revision Number
- G. Doc File Extension

Tips for creating your own FNC:
- Identify 4-7 important elements about the data. Elements could include date (YMMDD format), creator, location, project name, etc.
- Separate elements with underscores (file_name.doc) or dashes (file-name.doc). Some software doesn’t recognize file names with spaces.
- Avoid special characters such as ~ ! @ # $ % ^ & * ( ) ` ; < > ? , [ ] { } ' “ and |.
- Strike a balance: too few elements creates ambiguity, too many elements limits discoverability.
- Err on the side of brevity by creating meaningful abbreviations.
- Create documentation on your FNC that includes elements used, abbreviations used, updates/changes to the FNC, etc.
- The order of the elements should be from general to specific.
- File versioning allows you track the progress of your work. An easy way to track file versions is to add 'v01', 'v02', etc. at the end of your FNC. Don't use ambiguous version names such as 'final', 'updated', 'new', etc.

PHYSICAL DATA DOCUMENTATION & ORGANIZATION

For physical data, it’s best to create documentation and organization procedures before collecting data because it can be difficult and time consuming to reorganize in the middle of a research project. You can apply a file naming convention to physical data in order to create a consistent labeling system and then create documentation explaining the physical layout of the data.

OU LIBRARIES CAN HELP WITH DATA DOCUMENTATION & ORGANIZATION

Contact Joanna Thielen, Research Data Librarian, for assistance
jthielen@oakland.edu | 248-370-2477

Source: Data Management for Researchers by Kristin Briney
DATA SHARING

WHAT IS DATA SHARING?

Sharing your data makes it available to other researchers to re-use (collaborators, other researchers, the general public, etc.). Data sharing is increasingly becoming mandated by institutions, funding agencies and publishers.

WHY IS DATA SHARING IMPORTANT?

- Increases research impact and citation rate
- Promotes collaboration
- Increases transparency and reproducibility of research
- May be mandated by funding agency or publisher

BEST PRACTICES

- Check if your funding agency has data sharing requirements
  - They may require that the data must be made available for a certain period of time
  - They may also require using specific data repositories or that you provide them with a citation for the dataset

- Select data to share
  - You should share the data that supports all parts of your publications, including tables and figures
  - However, confidential data should not be shared

- Share data around the time of publication
  - Some funders and publishers require that the data is available prior to publication

- Apply a license to your data to facilitate re-use
  - Apply a Creative Commons license (creativecommons.org/licenses/) or an Open Data Commons license (opendatacommons.org/licenses/)
  - The license will tell other researchers how they can (and can't) re-use your data

- Before sharing, check intellectual property rights
  - Data itself is considered a fact and isn't subject to copyright
  - But an expression of the data (such as a table or figure) can be subject to copyright

Source: Data Management for Researchers by Kristin Briney
HOW TO SHARE YOUR DATA

There are several options for how to share your data. Here are four common options:

1. **Submit data as supplemental files to a publication** – Check the journal’s Author Guidelines to see if this is an option.

2. **Deposit it in a data repository** – A data repository makes your data publicly available online. Most repositories give a Digital Object Identifier (DOI) to the deposited dataset, which is a unique identifier that other researchers can use to cite your dataset. Examples of data repositories:
   - **Dryad**
     - Disciplines: Sciences & medicine
     - URL: dryad.org
   - **ICPSR**
     - Disciplines: Social Sciences
     - URL: icpsr.umich.edu
   - **UK Data Archive**
     - Disciplines: Social sciences & humanities
     - URL: data-archive.ac.uk
   - **Figshare**
     - Disciplines: All disciplines
     - URL: figshare.com

How to find a data repository for your data:
   - Registry of Research Data Repositories (re3data.org)
   - Simmons College list of data repositories by discipline (bit.ly/data_repositories)

3. **Publish it in a data journal** – As the name implies, a data journal publishes the actual data files along with a description of the dataset. Examples include *Scientific Data* (published by Nature) and *Research Data Journal for the Humanities and Social Sciences* (see right).

4. **Personal website** – This option is not recommended due to the significant time commitment required to maintain the website.

SHARING CONFIDENTIAL DATA

According to OU Policy 860, confidential research data includes (but is not limited to) “information related to a forthcoming or pending patent application, grant applications and proposals, and information related to human subjects.” Confidential data should not be shared publicly. Deidentified confidential data possibly can be shared (in case of research data that contains human subject information).

OU LIBRARIES CAN HELP WITH DATA SHARING

Contact Joanna Thiel, Research Data Librarian, for assistance
jthielen@oakland.edu | 248-370-2477

Source: *Data Management for Researchers* by Kristin Briney
DATA PRESERVATION

WHAT IS DATA PRESERVATION?

Data preservation is the series of managed activities necessary to ensure continued access to your research data for as long as necessary. These activities are undertaken after the data is no longer in active use.

WHY IS DATA PRESERVATION IMPORTANT?

You worked hard to gather your research data. Taking steps to preserve your data will ensure that you (and others) will be able to access and re-use it at a later date. Unlike most physical data, digital data needs to be carefully curated to mitigate the risk of software obsolescence, hardware failure, etc.

BEST PRACTICES

Select data to retain

- It's best to err on the side of keeping more data than less
- But you probably won't have the storage capacity to keep all of your data
- Always keep data that is used in publications, supports intellectual property claims, or can't be easily reproduced

Convert your data to non-proprietary formats

- Proprietary file formats and software become obsolete quickly
- Choosing non-proprietary file formats (CSV, PDF, MP3, etc.) will help to provide access to your data over the long term
- Stanford Library has an extensive list of suggested non-proprietary file formats: bit.ly/file_formats

Retain your data for 5-10 years

- Your funding agency may have their own data retention policy
- The minimum amount of time to retain data from federally funded research projects is 3 years
- Review your data annually and remove unneeded files

Source: Data Management for Researchers by Kristin Briney
HOW TO PRESERVE YOUR DATA

There are several options for how to preserve your data. Here are two common options:

5. **Preserving your data yourself** – You can preserve your data using your own hardware (such as an external hard drive). If you choose this option, your preserved data should be backed up (see Data Storage & Security Best Practices handout for more info). The major disadvantage of self-preservation of data is the significant time investment required.

6. **Using a data repository** – A data repository is an online storage location for research data, usually for a specific discipline. Most data repositories make deposited data publicly available, making them an ideal way to share your data. **Before depositing your data,** carefully read the repository’s Terms of Service to determine if they provide long-term preservation. Examples of data repositories:

![Dryad](dryad.org)
Disciplines: Sciences & medicine
URL: dryad.org

![ICPSR](icpsr.umich.edu)
Disciplines: Social Sciences
URL: icpsr.umich.edu

![UK Data Archive](data-archive.ac.uk)
Disciplines: Social sciences & humanities
URL: data-archive.ac.uk

![Figshare](figshare.com)
Disciplines: All disciplines
URL: figshare.com

You can find other data repositories using the Registry of Research Data Repositories (re3data.org) or Simmons College list of data repositories by discipline (bit.ly/data_repositories).

PRESERVING & DESTROYING CONFIDENTIAL DATA

According to OU Policy 860, confidential research data includes (but is not limited to) “information related to a forthcoming or pending patent application, grant applications and proposals, and information related to human subjects.”

**Preservation**
- Preserve in a secure, controlled-access environment
- De-identified data possibly can be deposited into a data repository

**Destruction**
- Work with departmental IT to ensure that digital files containing confidential data are deleted from your storage media if no longer needed

OU LIBRARIES CAN HELP WITH DATA PRESERVATION

**Contact Joanna Thielen, Research Data Librarian, for assistance**

jthielen@oakland.edu | 248-370-2477

Source: *Data Management for Researchers* by Kristin Briney